Short Read Workshop Day 4
Trimming, Mapping, IGV

Lynn Sanford and Georgia Barone

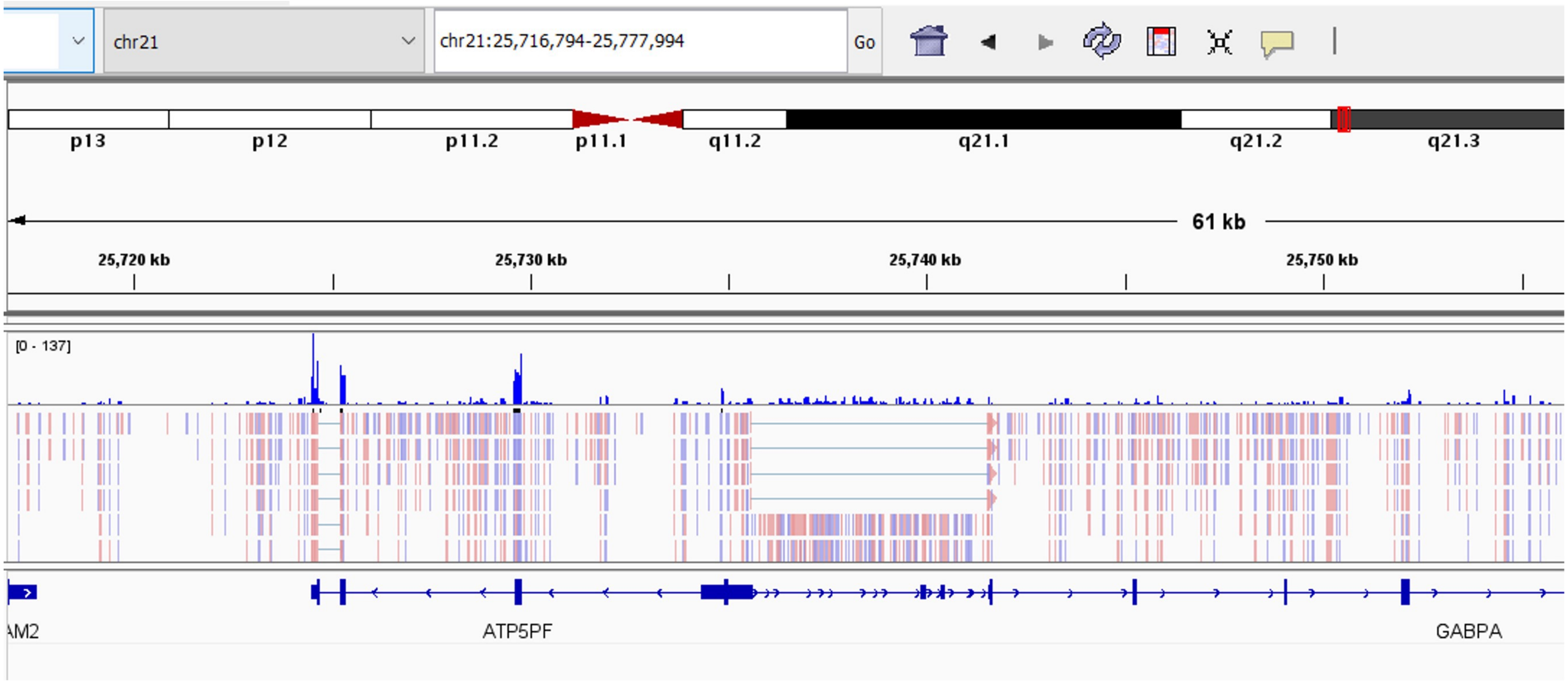
2024

Day 4 overview

- Trimming fastq files
- Mapping fastq files
- More about mapped file formats
- Visualizing mapped files

Goal of the Day

View sequencing data as reads aligned to a genome



Trimming/mapping recap

50 base read:
TAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGACCAGGATATAG

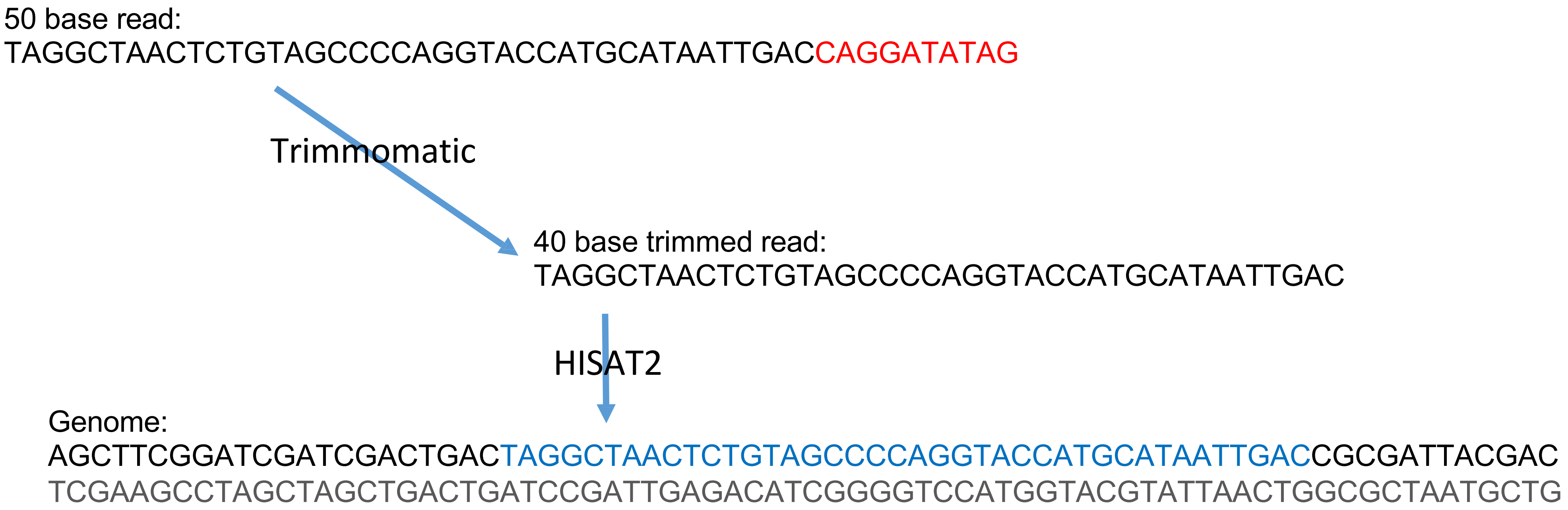
Trimming/mapping recap

50 base read:
TAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGACCAGGATATAG

Trimmomatic

40 base trimmed read:
TAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGAC

Trimming/mapping recap



50 base read:
TAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGACCAGGATATAG

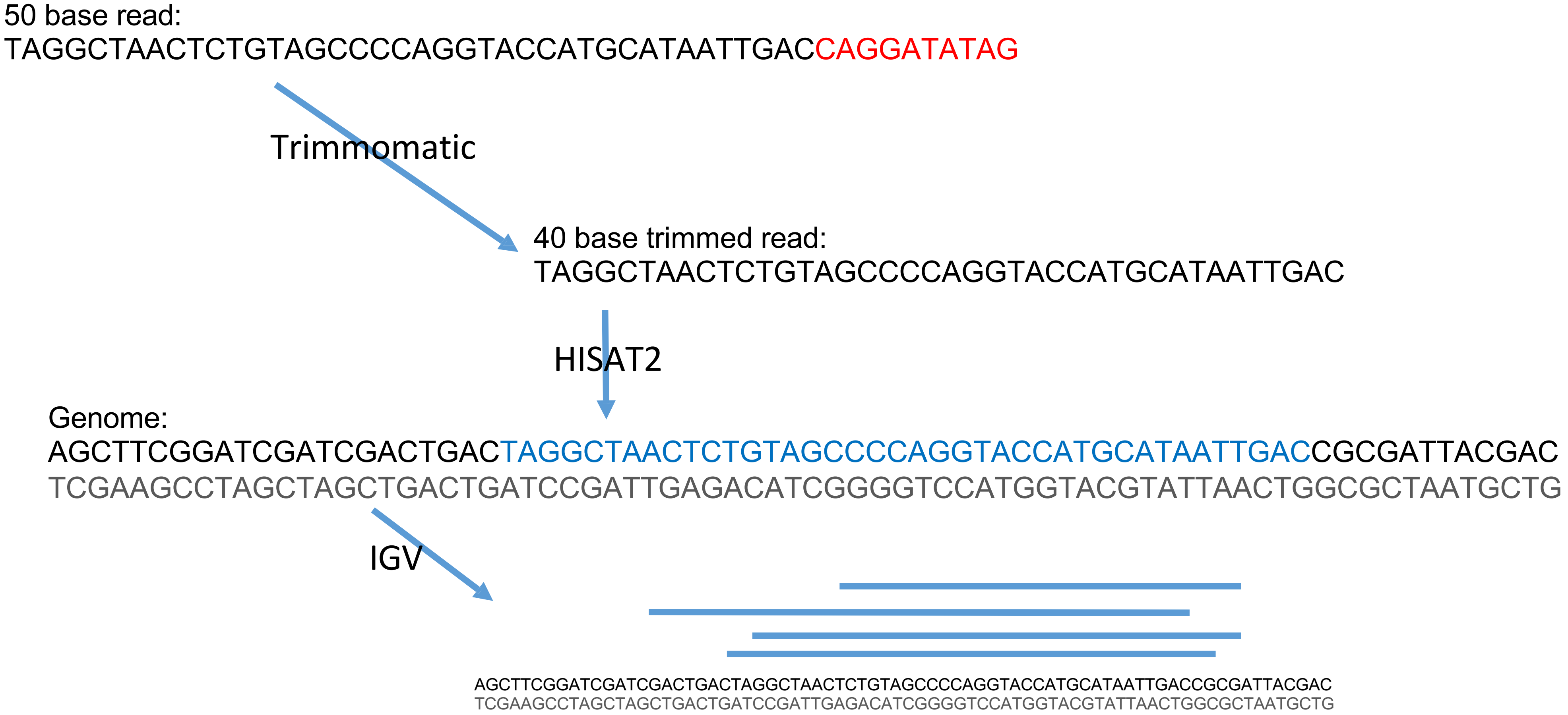
Trimmomatic

40 base trimmed read:
TAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGAC

HISAT2

Genome:
AGCTTCGGATCGATCGACTGACTAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGACCGCGATTACGAC
TCGAAGCCTAGCTAGCTGACTGATCCGATTGAGACATCGGGGTCCATGGTACGTATTAACTGGCGCTAATGCTG

Trimming/mapping recap



50 base read:
TAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGACCAGGATATAG

Trimmomatic

40 base trimmed read:
TAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGAC

HISAT2

Genome:
AGCTTCGGATCGATCGACTGACTAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGACCGCGATTACGAC
TCGAAGCCTAGCTAGCTGACTGATCCGATTGAGACATCGGGGTCCATGGTACGTATTAACTGGCGCTAATGCTG

IGV

AGCTTCGGATCGATCGACTGACTAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGACCGCGATTACGAC
TCGAAGCCTAGCTAGCTGACTGATCCGATTGAGACATCGGGGTCCATGGTACGTATTAACTGGCGCTAATGCTG

Trimming fastq files with Trimmomatic

- Follow Variables/Trimmomatic worksheets to:
- Learn more about variables in bash
- Create Day4 directories
- Edit script to run Trimmomatic

  Input: fastq files with full-length reads

  Output: fastq files with trimmed/filtered reads

- Extra: Edit the d4_trim_qc.sbatch script to run pre-trim qc and compare the plots/stats to post-trim

How do you trim polyA regions from both sides of reads?

How do you trim polyA regions from both sides of reads?

- Make a new fasta file with a polyA segment, or append to the Illumina adapter file, if writeable
- \>polyA

  AAAAAAAAAAAAAAAAAAAAAAAA
- ILLUMINACLIP:<new fasta file>:2:30:10

Mapping fastq files with HISat2

- Follow Mapping/IGV worksheet to:
- Rsync the mapping script
- Edit script and run HISAT2

  Input: trimmed fastq files

  Outputs: .sam, .bam, .sorted.bam, .bam.bai files
- Visualize BAM file on the IGV web app

Homework

Day 4 Homework – FASTQC, trimming, mapping, IGV

The assessment tomorrow will run many of the same steps as this homework. These steps are essential in ALL short read data processing.

Variables – evaluating (calling)

$variablename

```
echo $a
```

```
grep $gene_name <filename>
```

```
trim_script=“$filepath”/d4_trim_qc.sbatch
OR
trim_script=${filepath}/d4_trim_qc.sbatch
```

```
wc $filelist
```

Several ways of evaluating:

```
$a

${a}

“$a”
```

These differ slightly, and you will see us use them all in scripts